Predicting Pneumonitis

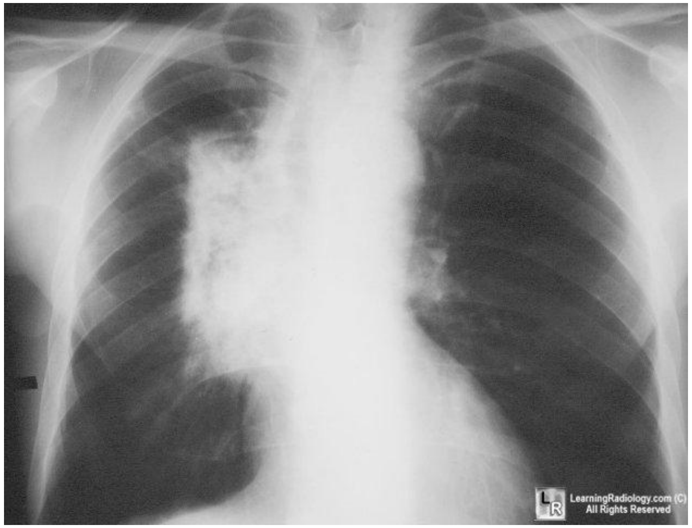
Improving clinical relevance in ensemble support vector machine models of radiation pneumonitis risk

Todd W. Schiller, Yixin Chen
Washington University in St. Louis
Department of Computer Science

Issam El Naqa, Joseph O. Deasy
Washington University in St. Louis
School of Medicine

A Potentially Fatal Problem

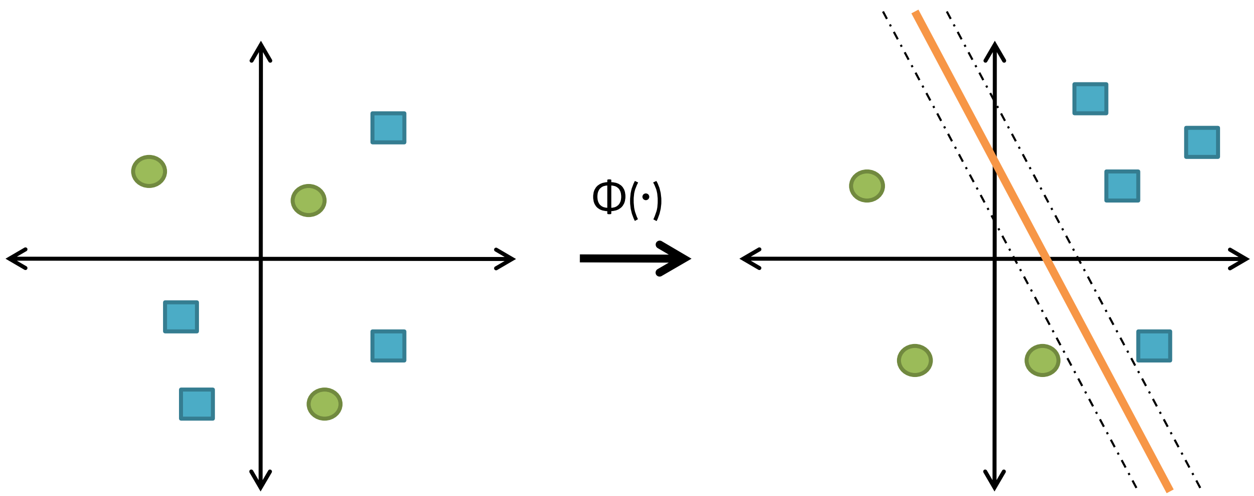
- Inflammation of the lungs that can result from thoracic radiation therapy
- Advanced predictive models are binary-outcome
- Patient risk is not black or white



From learningradiology.com

An Imperfect Solution

- Combine output from a set of support vector machines (SVMs)



An Imperfect Solution

- Each SVM has its own set of features; does not identify core factors affecting risk
- Ensemble SVM is still binary-outcome

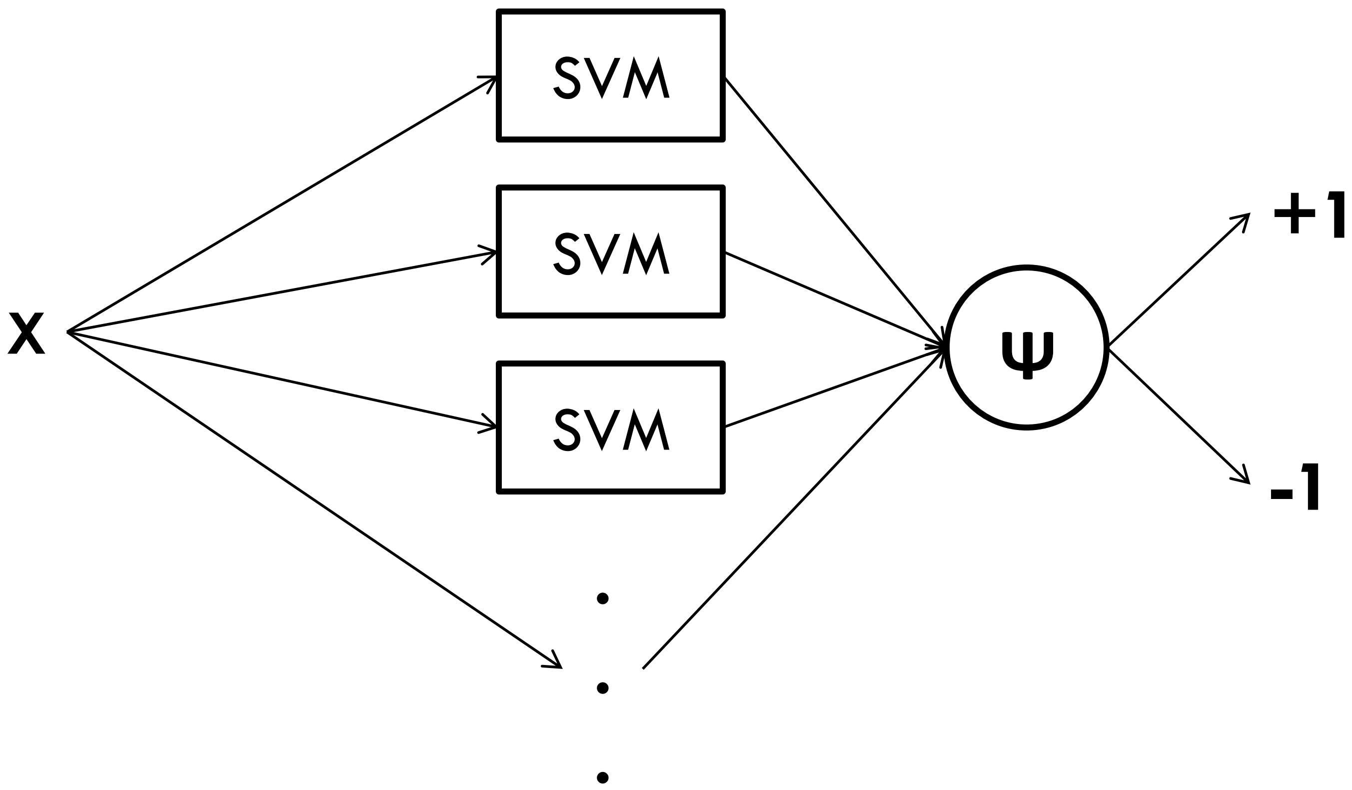
Our Solution

- **Parsimony:** introduce unification step to feature selection
- **Performance:** larger ensembles provide statistically significant benefits to AUC
- **Probabilistic Output:** extend model to produce risk estimates

6

Parsimony

SVM Ensemble



From Schiller et. al

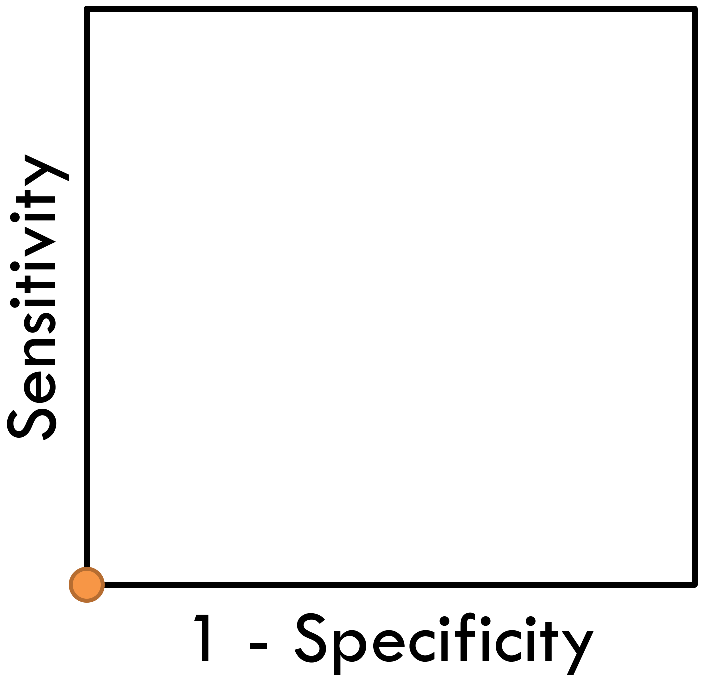
Unified Feature Set

- Create a separate set of feature selection SVMs for each fold
- Perform Chen et al.'s algorithm on each SVM to pick at most 5 features (maximizing AUC)
- Count the number of times each feature has been selected by one of the feature selection SVMs

Most Commonly Selected Features Across LOO Folds
COMLAT
COMSI
Performance Status
Max. Dose to the Heart

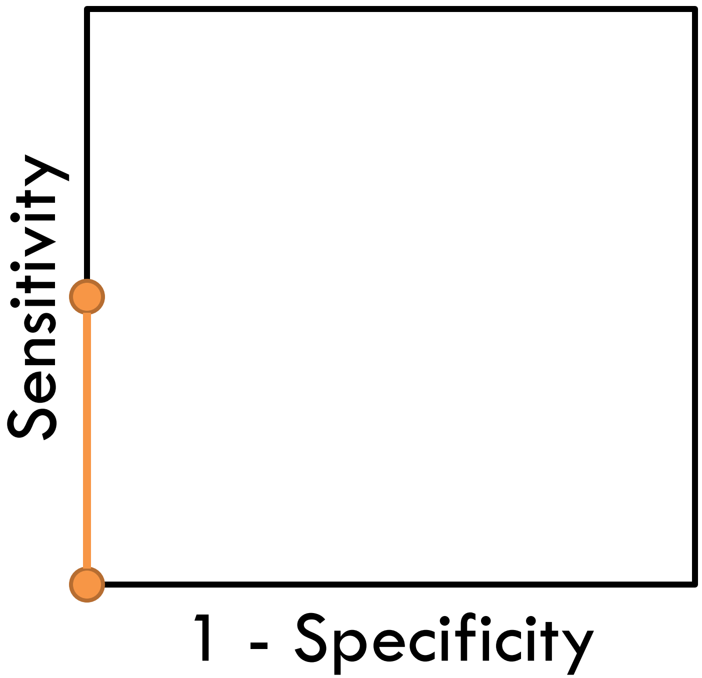
Performance

Receiver Operating Characteristic (ROC) Curve



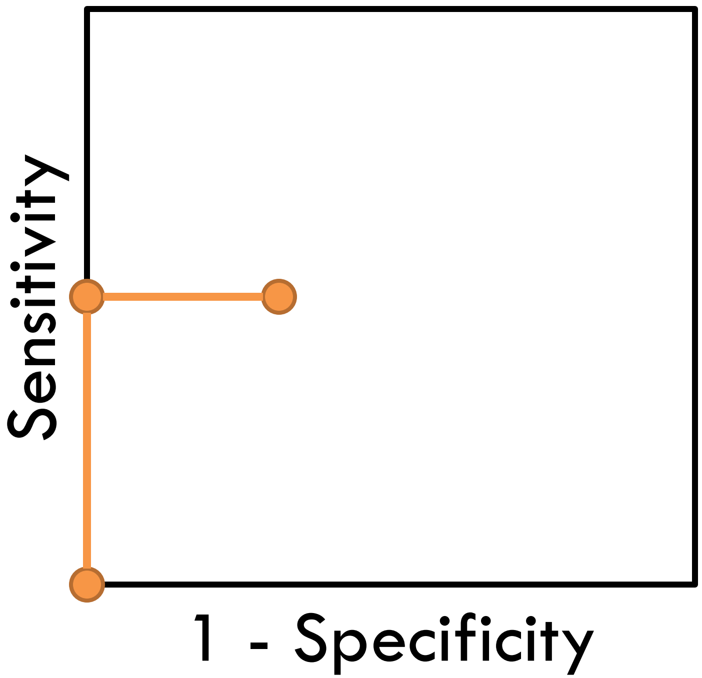
Patient	Score	Actual
A	.75	+1
E	.30	-1
B	.15	+1
D	-.15	-1
C	-.50	-1

Receiver Operating Characteristic (ROC) Curve



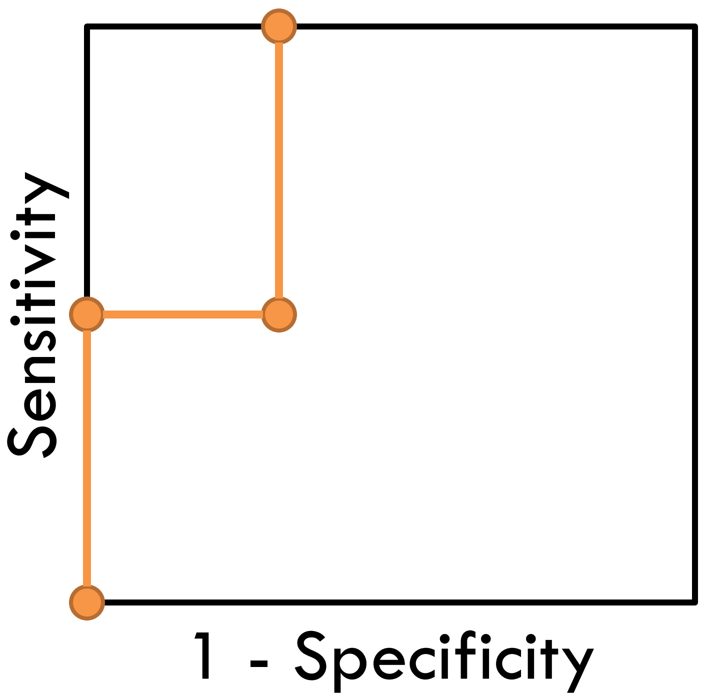
Patient	Score	Actual
A	.75	+1
E	.30	-1
B	.15	+1
D	-.15	-1
C	-.50	-1

Receiver Operating Characteristic (ROC) Curve



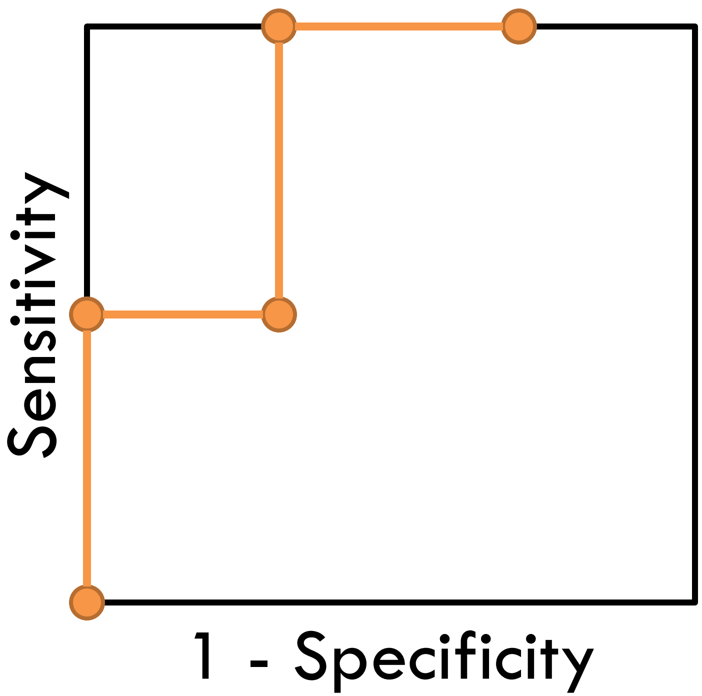
Patient	Score	Actual
A	.75	+1
E	.30	-1
B	.15	+1
D	-.15	-1
C	-.50	-1

Receiver Operating Characteristic (ROC) Curve



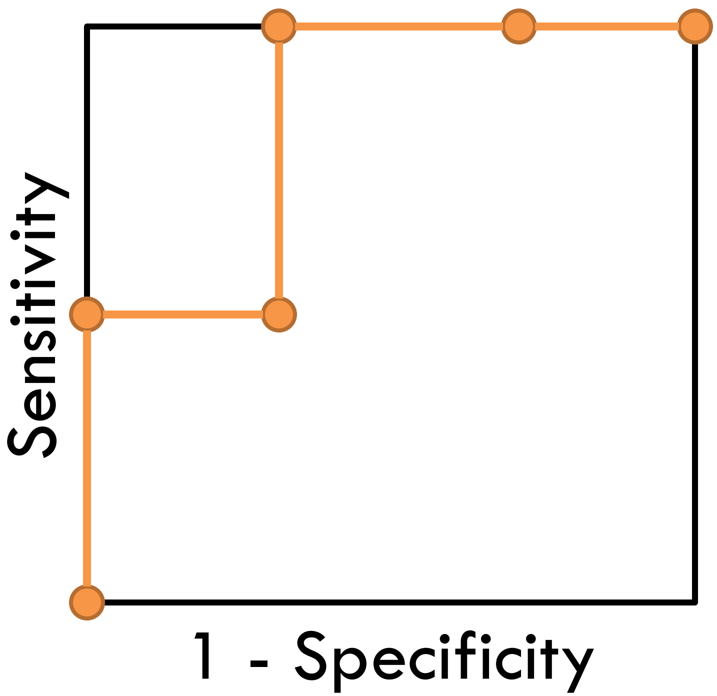
Patient	Score	Actual
A	.75	+1
E	.30	-1
B	.15	+1
D	-.15	-1
C	-.50	-1

Receiver Operating Characteristic (ROC) Curve



Patient	Score	Actual
A	.75	+1
E	.30	-1
B	.15	+1
D	-.15	-1
C	-.50	-1

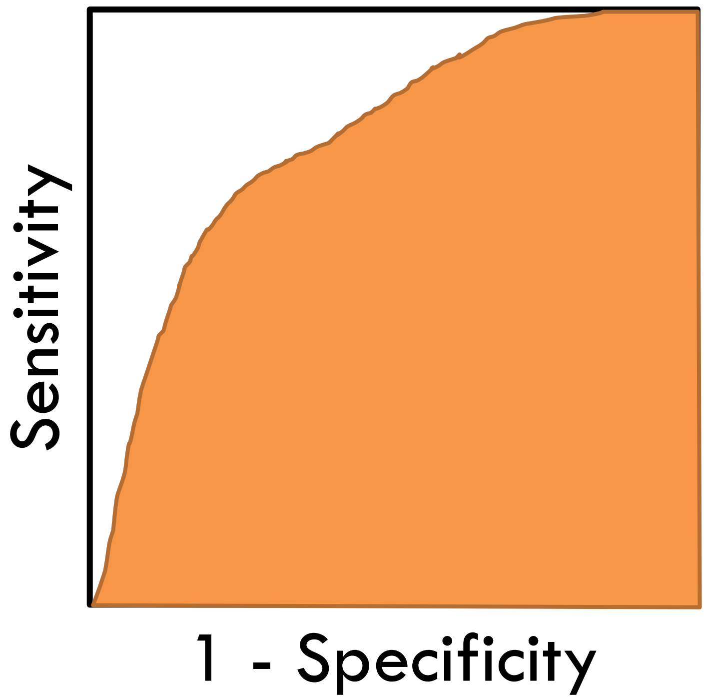
Receiver Operating Characteristic (ROC) Curve



Patient	Score	Actual
A	.75	+1
E	.30	-1
B	.15	+1
D	-.15	-1
C	-.50	-1

Area under the ROC

- A metric of model performance
- Probability that a random positive instance will receive a higher score than a random negative instance



Testing Ensemble Performance

- Compare paired differences in 10-fold AUCs for ensembles with n=1,3,5,10 component SVMs
- 100 trials (different foldings)

n	Min. AUC	Mean AUC	Max. AUC
1	0.5828	0.6959	0.7712
3	0.6486	0.7246	0.7853
5	0.6786	0.7374	0.7940
10	0.6925	0.7501	0.7937

Testing Ensemble Performance

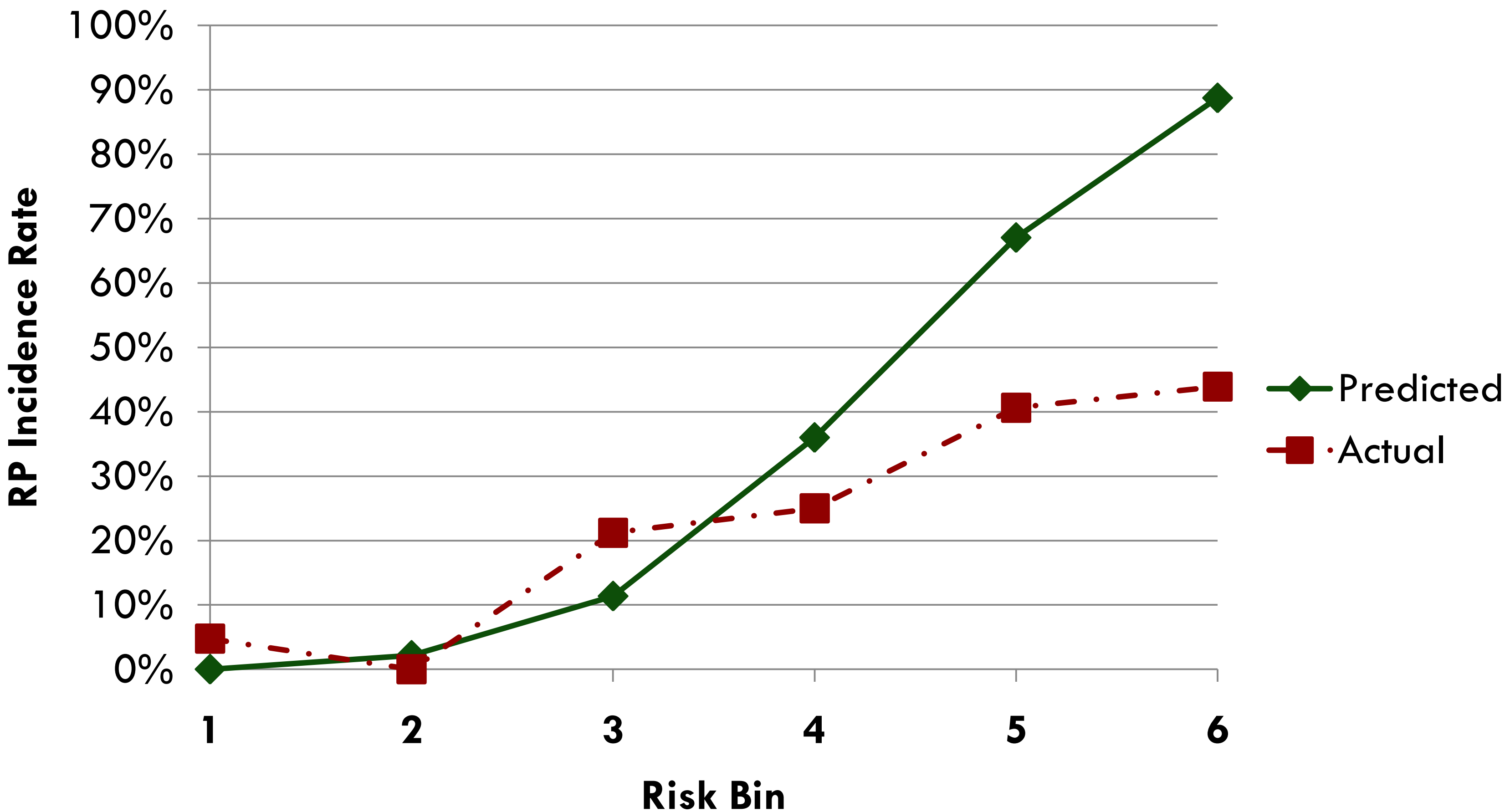
- Jarque-Bera tests only reject normality for n=5 v. n=1 case (5% significance level)
- One-tailed paired t-test indicates **statistically significant improvement in AUC** for specified increases in ensemble size (5% significance level)

Impacts

- Statistical evidence of ensemble method benefits
- Multiple types of classifiers aren’t necessary for synergy

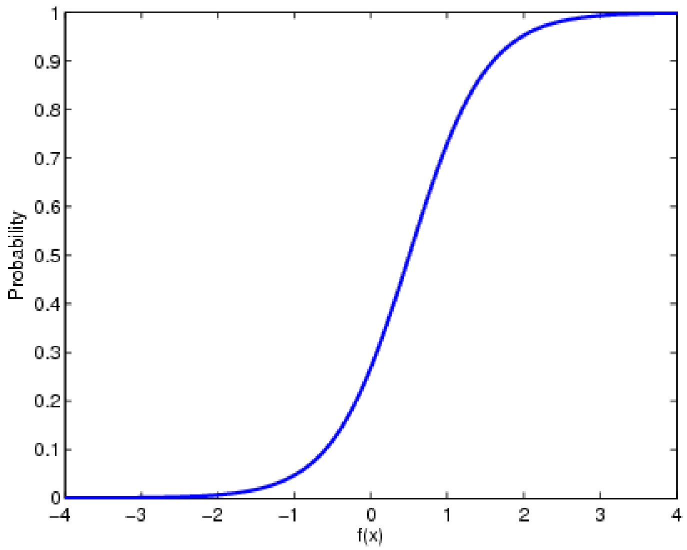
Probabilistic Output

Binary-Averaging Risk Binning

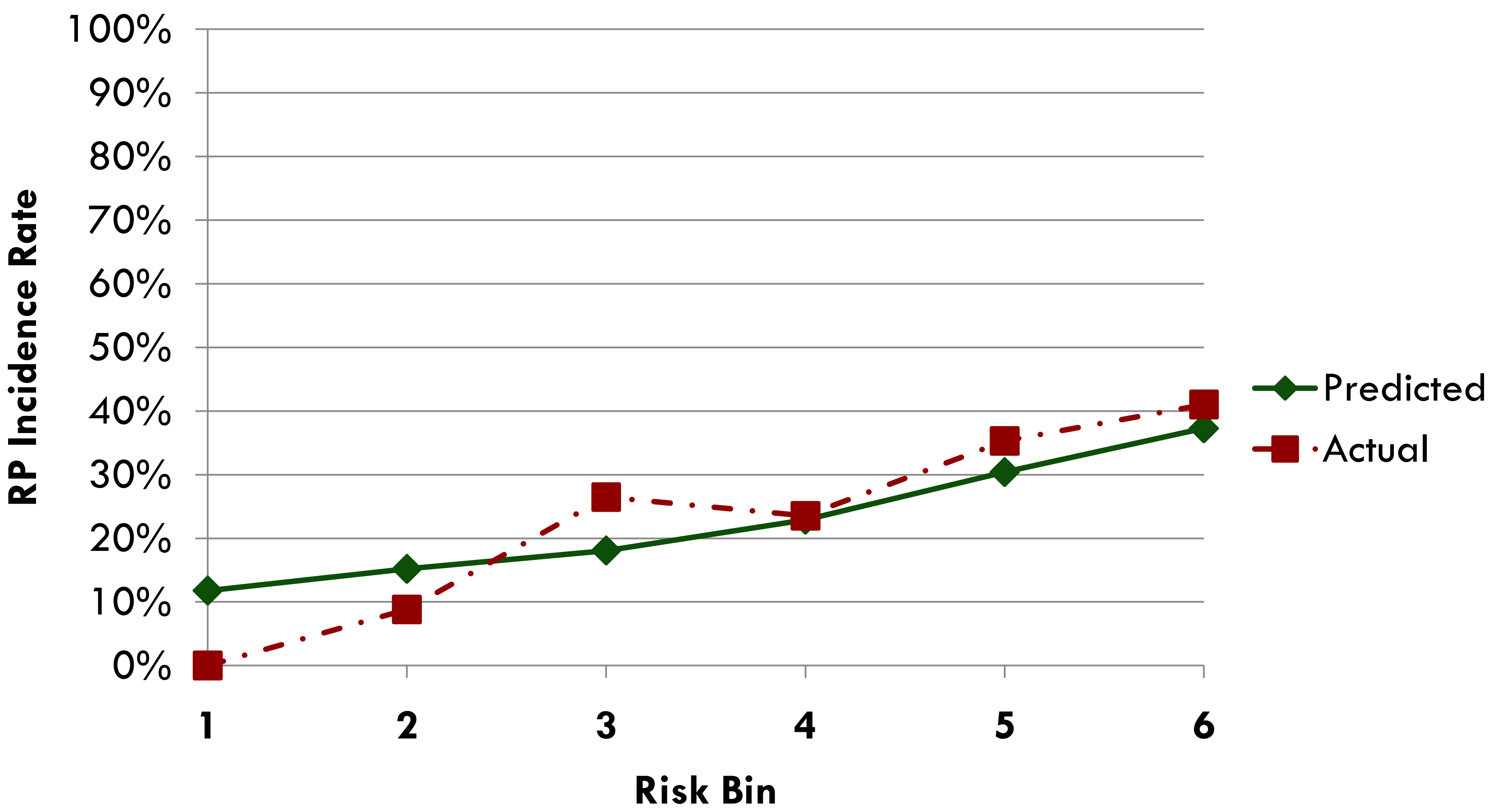


Probabilistic Tuning

- John Platt's idea: fit the output of the SVM to a sigmoid curve with parameters *A* and *B*
- Pick *A* and *B* to minimize the cross-entropy error



Probabilistic Risk Binning



Impacts

- Introduce probabilistic tuning to the RP domain
- Predicted RP Incidence
  - Probability is understood by patients
  - Assess effect of potential treatment change on risk
- Actual RP Incidence / Binning
  - Highlights importance of model AUC
  - Shows binary-averaging is not a proper proxy for probability

25

Conclusion

Our Solution

- **Parsimony:** introduce unification step to feature selection
- **Performance:** larger ensembles provide statistically significant benefits to AUC
- **Probabilistic Output:** extend model to produce risk estimates

Directions for Future Research

- **Binning** - what comes first: better AUCs or better probability estimates?
- **Increasing AUC** – creating composite features to leverage domain knowledge